Assaying chromosomal inversions by single-molecule haplotyping

Daniel J Turner[1], Jay Shendure[2], Greg Porreca[2], George Church[2], Peter Green[3], Chris Tyler-Smith[1] & Matthew E Hurles[1]

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

Inversions are an important form of structural variation, but they are difficult to characterize, as their breakpoints often fall within inverted repeats. We have developed a method called 'haplotype fusion' in which an inversion breakpoint is genotyped by performing fusion PCR on single molecules of human genomic DNA. Fusing single-copy sequences bracketing an inversion breakpoint generates orientation-specific PCR products, exemplified by a genotyping assay for the *int22* hemophilia A inversion on Xq28. Furthermore, we demonstrated that inversion events with breakpoints embedded within long (>100 kb) inverted repeats can be genotyped by haplotype-fusion PCR followed by bead-based single-molecule haplotyping on repeat-specific markers bracketing the inversion breakpoint. We illustrate this method by genotyping a Yp paracentric inversion sponsored by >300-kb-long inverted repeats. The generality of our methods to survey for, and genotype chromosomal inversions should help our understanding of the contribution of inversions to genomic variation, inherited diseases and cancer.

Structural variation within the human genome (deletions, duplications, inversions and translocations) accounts for an appreciable proportion of both causative alleles in genetic disorders with simple inheritance patterns, and susceptibility alleles for infectious diseases[1,2]. Therefore, it is to be expected that an as-yet unknown proportion of the genetic risk in complex diseases will also be determined by structural variants. Current methods for detecting variants conferring disease susceptibility are biased against identifying structural variants[1]. To address this deficit, comprehensive surveys of structural variation are required, allied to assays that allow population-scale genotyping in association studies of the variants discovered. Population-scale genotyping of structural variation is also necessary to investigate the role that structural changes have had in recent human evolution[3].

Information on the locations and frequency of copy-number polymorphisms in the human genome is accruing rapidly, largely from microarray-based comparative genome hybridization studies of increasing resolution[4–6]. In addition, diverse quantitative methods allow population-scale multiplexed genotyping of copy-number variants[7].

By contrast, in the absence of a similarly high-throughput discovery method for inversions, fewer inversion polymorphisms have been identified. Although recent analysis of dense fosmid end sequencing data from a single individual has been used to identify new inversion variants on a genome-wide basis[8], this method does not easily scale to the analysis of the tens of samples needed to identify all common (frequency greater than 5%) inversions. Similarly, there are no genotyping methods that allow population-scale genotyping of known inversion variants: the usefulness of single-nucleotide polymorphisms (SNPs) in linkage disequilibrium with an inversion is possibly limited by multiple independent origins for the same inversion[9,10]. Known inversions are typically assayed singly by laborious interphase fluorescence *in situ* hybridization or pulsed-field gel electrophoresis[11], neither of which is appropriate for population-scale genotyping. A handful of well-characterized inversions can be genotyped singly by long PCR or restriction fragment length polymorphism (RFLP) analysis using standard agarose gel electrophoresis[12,13], but this is only practical when inversion breakpoints are not embedded within long (>10 kb) inverted repeats. Unfortunately, among the different classes of structural variants, inversions are the most likely to feature rearrangement breakpoints within long, highly similar duplicated sequences[8].

To address these deficits, we have developed PCR-based single-molecule haplotyping methods that enable both surveys for new inversion variants and population-scale genotyping of known inversions. We demonstrate our ability to haplotype single DNA molecules that span known or likely inversion breakpoints by developing a PCR-based genotyping assay for the hemophilia-causing inversion that disrupts the coagulation factor VIII (*F8*) gene on Xq28 (ref. 14). Moreover, we demonstrate that these methods can be used to genotype known inversions with breakpoints embedded in inverted repeats hundreds of kilobases long by developing an assay for the Yp inversion that influences the germline rate of infertility-causing XY translocations[15].

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Medical and Molecular Genetics, King's College School of Medicine, King's College, London, WC2R 2LS, UK. Correspondence should be addressed to M.E.H. (meh@sanger.ac.uk).

RECEIVED 20 DECEMBER 2005; ACCEPTED 14 APRIL 2006; PUBLISHED ONLINE 23 MAY 2006; DOI:10.1038/NMETH881

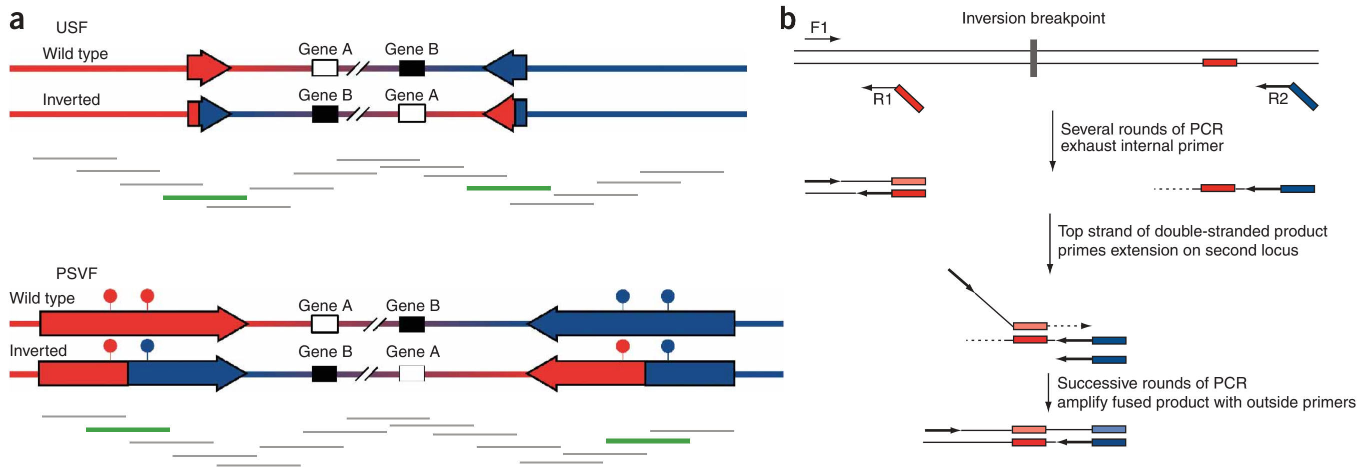

Figure 1 | Strategies for using haplotype-fusion PCR for genotyping inversions. (**a**) In the two inversion cases shown, the inversion breakpoints lie within inverted repeats (red and blue arrows). In the first case (unique sequence fusion, USF; top), the inverted repeats are shorter than the size of the DNA fragments in the genomic DNA preparation (gray and green), whereas in the second case (paralogous sequence variant fusion, PSVF; bottom) the inverted repeats are longer than the DNA fragments. DNA fragments that are informative for genotyping the inversion are shown in green. The filled circles above the inverted repeats in the second inversion represent paralogous sequence variants that can be used to distinguish between the two repeats, the haplotype of which is diagnostic for the state of the inversion. (**b**) A double-stranded DNA at either side of an inversion breakpoint (top), is amplified by three primers (black arrows) in a PCR conducted in an oil:water emulsion. The red tag on the internal primer R1 is an oligonucleotide that is identical to the sequence on the lower strand of the right-hand amplicon. The blue tag on the right-hand outside primer R2, is an oligonucleotide sequence (Zip1) that facilitates subsequent bead-based haplotyping. In the early rounds of PCR, a double-stranded product encompassing the left-hand amplicon (F1–R1) is amplified exponentially, and a single-stranded product is synthesized linearly from the R2 primer. After the internal primer R1 is exhausted, the upper strand of the left-hand amplicon acts as a forward primer for amplification of the right-hand amplicon.

RESULTS

Inversion genotyping

The inversion genotype of an individual can be deduced from their genomic DNA if single-molecule haplotyping is performed on the DNA fragments that span the inversion breakpoints (**Fig. 1a**).

Fusion PCR[16,17] performed on single-molecule templates can juxtapose sequences that lie on either side of an inversion breakpoint. The haplotypes of these sequences are diagnostic for the inversion. Fusing the sequences allows us to apply a greater range of downstream haplotyping techniques. PCR in oil:water emulsions[18] provides a means to amplify single-molecule templates on a far larger scale (10^3–10^4-fold greater) than performing single-molecule dilutions in a microtiter plate, because amplification occurs in millions of aqueous compartments that are stable

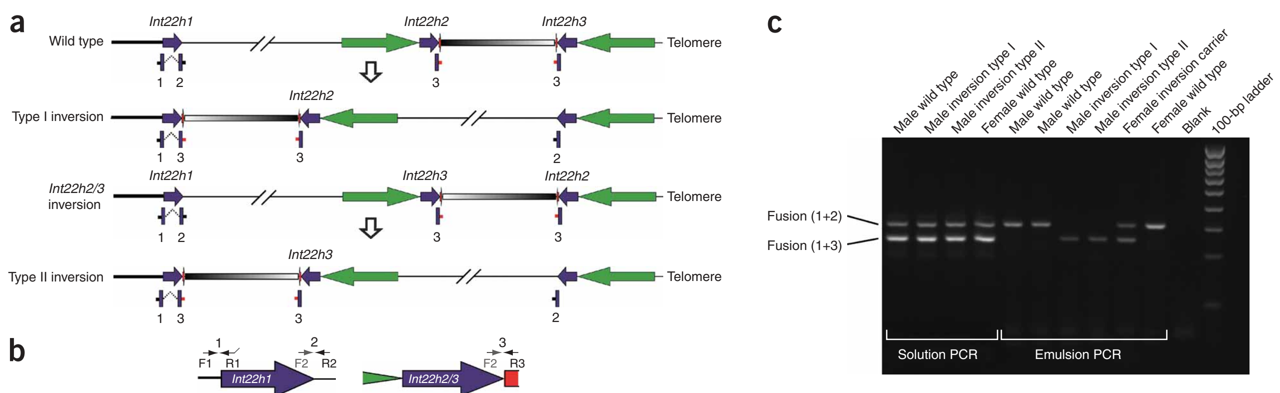

Figure 2 | Genotyping the hemophilia inversion. (**a**) The four possible orientations of the genomic region on Xq28 containing the three *int22h* repeats (purple arrows). *Int22h2* and *int22h3* are embedded in longer (~50 kb) inverted repeats separated by ~70 kb of unique sequence (black-white gradient). The portions of this larger segmental duplication that are not found in the *int22h* repeats are shown as green and red arrows. The four possible orientations refer to the reference 'wild-type' sequence, the *int22h2/3* inverted sequence, and type I and type II inversions. Beneath each schematic are the three PCR products (labeled 1, 2 and 3) amplified using the primers described in the text. The two PCR products fused by haplotype-fusion PCR are linked by a dashed line. The red and black boxes represent the ends of the amplicons that do not lie within the *int22h* (purple) repeats. The red boxes indicate that the end of the amplicon lies within the repeated sequences shown in red. The black boxes indicate the end of the amplicon lies within a unique sequence. (**b**) The position of the PCR primers F1, R1, R2 and R3 on the *int22h* repeats. The arrow labeled F2 shows the binding site complementary to the tail of the R1 primer. (**c**) An agarose gel of the results of haplotype-fusion PCR when performed in solution and in emulsions on all possible inversion genotypes. The four inversion genotypes that produce single band patterns in emulsions all produce two band patterns when the fusion PCR is performed in solution because PCR products can be fused from different DNA templates.

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

to thermocycling[19]. The number of template molecules is far less than the number of compartments, and so amplification within a compartment derives from a single molecule. Therefore, we developed haplotype-fusion PCR to combine the advantages of fusion PCR and single-molecule amplification in emulsions (**Fig. 1b**).

In practice, the haplotyping method used to genotype the inversion differs depending on the length of the inverted repeat relative to the size of DNA fragments in the genomic DNA preparation (**Fig. 1a**). If the inverted repeat is shorter than the DNA fragment lengths, single-copy sequences on either side of an inverted repeat can be fused and the inversion genotyped by determining the identity of the fused PCR product. This is most easily exemplified by generating fused products of different sizes for the two orientations of the inversion. We term this method unique sequence fusion (USF).

If the inverted repeat is longer than the DNA fragments, then the inversion is genotyped by fusing two sequences that contain repeat-specific markers that lie on either side of the inversion breakpoint. Repeat-specific markers that are embedded within a long region of homology are known as paralogous sequence variants (PSVs). We term this method PSV fusion (PSVF).

Genotyping the *int22h* hemophilia inversion

The *int22h* inversion, which is a common cause of hemophilia A, is effected by *int22h* repeats that lie in inverted orientation on Xq28 (**Fig. 2a**). The proximal inverted repeat (*int22h1*) involved in the pathogenic inversion lies within the *F8* gene. There are two distal copies of the *int22h* repeat. At any one time, only one of these distal *int22h* repeats (*int22h2* and *int22h3*) lies in inverted orientation to *int22h1* (refs. 20,21). There are two major classes of hemophilia-causing inversions, type I and type II, and these differ in which of the two distal *int22h* repeats recombines with the *int22h1* repeat.

We designed a haplotype fusion–PCR assay to distinguish between the wild-type orientation of the hemophilia-causing inversion and the pathogenic orientation (either type I or type II). The F1 primer lies in the single-copy sequence proximal to the *int22h1* repeat. The R1 primer lies just inside the tail of the *int22h1* repeat. Thus amplicon 1, amplified by F1 and R1 primers, spans the boundary between the single-copy and the duplicated sequence. The tail of the R1 primer is complementary to a binding site just inside the head of the *int22h* repeat. We designed two other reverse primers: R2 lies in the single-copy sequence distal to the head of the *int22h1* repeat and leads to amplification of amplicon 2; R3 lies within the larger distal segmental duplication just adjacent to the heads of the *int22h2* and *int22h3* repeats and leads to amplification of amplicon 3 (**Fig. 2b**). The fusion of amplicon 1 and 2 during emulsion PCR is characteristic of the wild-type orientation, whereas the fusion of amplicons 1 and 3 generates a shorter PCR product that is diagnostic for the inverted orientation. When these fusion PCRs are performed in solution under standard PCR conditions both fusion products are generated in all individuals (**Fig. 2c**). However, when the fusion PCR is performed in emulsions, products from different templates can no longer fuse and only the fusion products expected for that specific inversion genotype are generated. Thus, this assay distinguishes between males with and without the inversion, female carriers of the inversion, and females homozygous for the wild-type orientation (**Fig. 2c**). To demonstrate the general utility of this assay, we genotyped (in blind setup) a mixed population of 10 men with and without the *int22h* inversion with 100% concordance with the known inversion genotypes (**Supplementary Fig. 1**).

Genotyping the Yp inversion by single-molecule haplotyping

The constitutive haploidy of the common 3.6-Mb Yp inversion[22] makes it an ideal test locus. This inversion has been suggested to result from homologous recombination between the ∼300-kb *IR3* inverted repeats[23]. It is known from previous studies that the

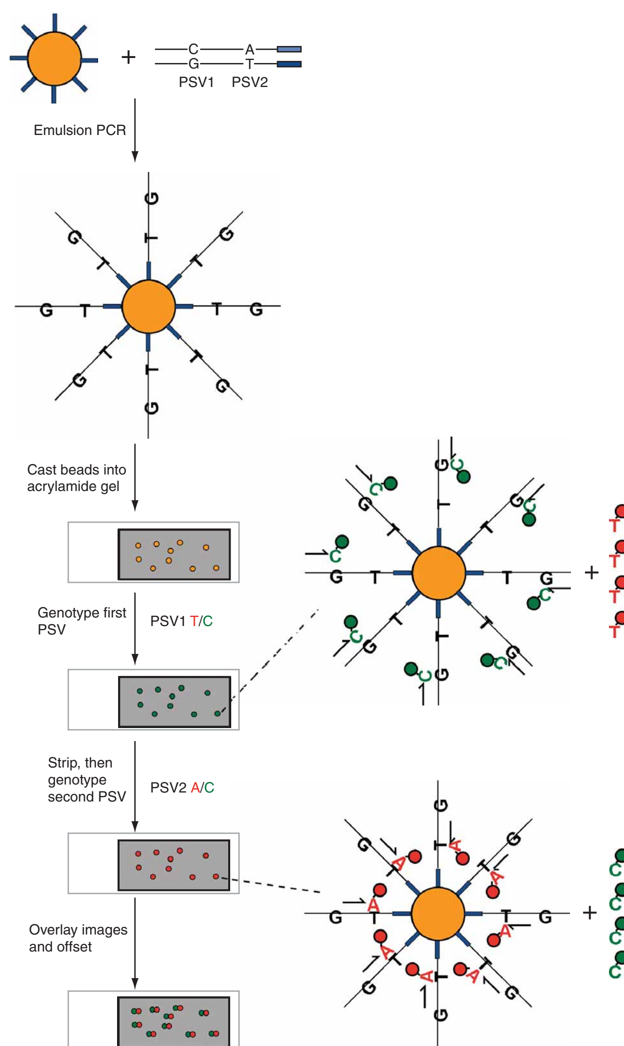

Figure 3 | Bead haplotyping. An emulsion PCR is performed in which the product of a haplotype-fusion PCR is amplified onto paramagnetic beads (orange) loaded with a universal oligonucleotide (blue) containing two repeat-specific markers (PSV1 and PSV2), the combination of which allows the inversion to be genotyped. For simplicity, only one double-stranded template labeled with the two bases to be genotyped is illustrated. Emulsion PCR generates beads coated in single-stranded amplified product from single-molecule templates. These beads are cast into a thin polyacrylamide gel (gray) on a microscope slide. Sequential rounds of genotyping by single-base extension with dye-labeled nucleotides (red and green circles) yield two images that can be overlaid and offset to reveal the haplotype on each bead. For each genotyping reaction, the annealing position of the single base extension primer (arrow) on the bead-immobilized template is shown, together with the dye-labeled nucleotide that is incorporated.

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

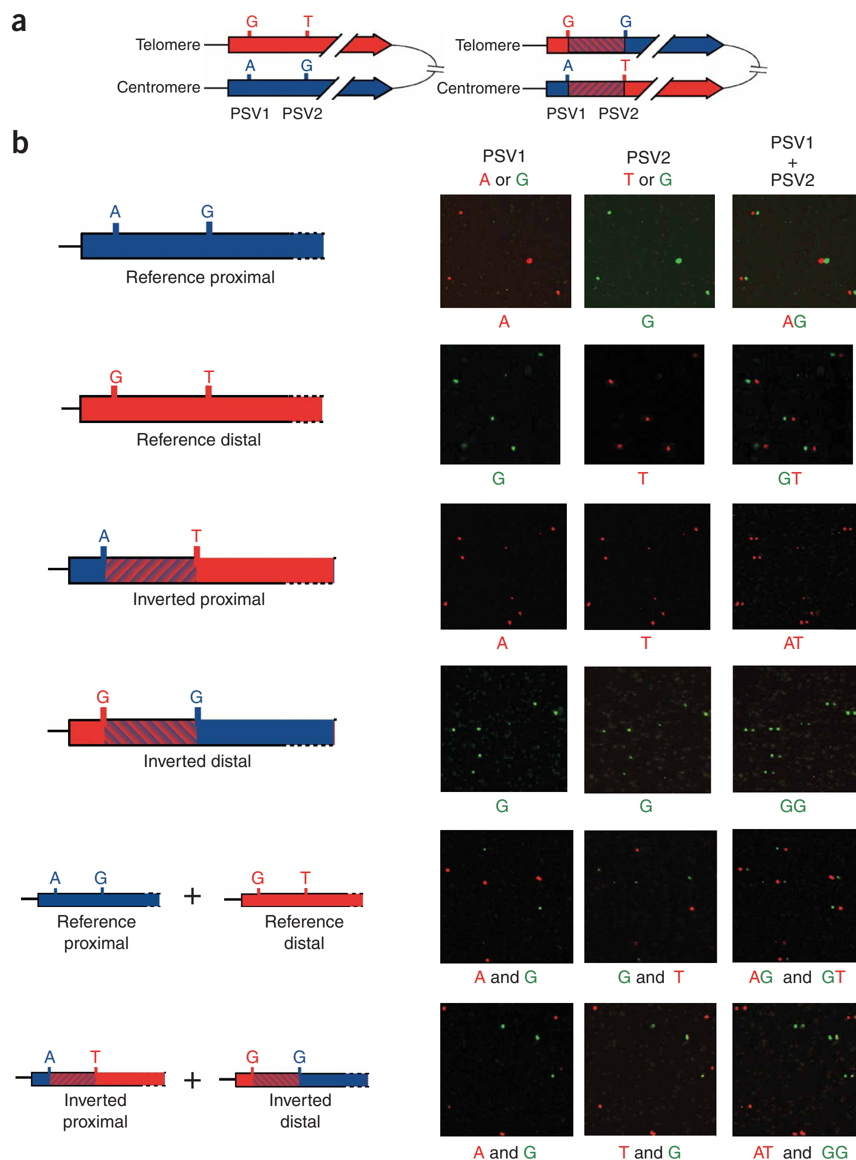

Figure 4 | Single-molecule haplotyping on YACs containing single *IR3* repeats. (**a**) The portions of the *IR3* repeats (red and blue arrows) encompassing the inversion breakpoint (striped interval) are aligned for both the reference (left) and inverted (right) orientations. Each *IR3* repeat is labeled with its respective haplotype of the two PSVs that flank the breakpoint. (**b**) The four possible PSV haplotypes apparent within the four *IR3*-containing YACs as well as the pair of PSV haplotypes obtained from mixing *IR3*-containing YACs from the reference and inverted orientations are illustrated (left). Each row shows the results of bead haplotyping conducted using the individual YACs and mixtures of YACs as templates for haplotype-fusion PCR and a subsequent haplotyping reaction. The bead haplotyping results are shown as three fluorescent microscopy images of the same portion of a larger image. The first one shows the genotype of the first PSV and the second, the genotype of the second PSV. The third image is the overlay of these two images, with the image of the genotype of the second PSV shifted 10 pixels to the right. The bases corresponding to the different colors are given above the images, and the genotype-haplotype observed is given below each image.

alternative orientations of the inversion (here denoted 'reference'—as found in the human reference sequence, and 'inverted') are present in different yeast artificial chromosome (YAC) libraries[24,25]. From these libraries, we identified 4 YACs (two from each orientation) that contain the *IR3*-proximal and -distal repeats and thus contain the inversion breakpoints. YAC clones *AA4A7/11* and *692B1* contain the proximal *IR3* repeat in 'inverted' and 'reference' orientations respectively. YAC clones *AMY8* and *821G7* contain the distal *IR3* repeat in 'inverted' and 'reference' orientations respectively. We localized the breakpoint of the Yp inversion within the ~300-kb-long *IR3* repeats to a ~7-kb interval bounded by two repeat-specific base substitutions, PSVs (manuscript in preparation; D.J.T., C.T.-S. and M.E.H.). Consequently, we can define haplotypes of these PSVs that are diagnostic of Y chromosomes having undergone this Yp inversion.

We used haplotype-fusion PCR to fuse two short amplicons containing the two PSVs defining the Yp inversion breakpoint interval. The two PSVs are separated by 50 bp in the fused ~300-bp product. The four YACs described above contain the four possible PSV haplotypes. We amplified fused haplotypes from these individual YACs, from pairwise mixtures of these YACs, and from genomic DNA of the same individuals used to generate 'reference' and 'inverted' YAC libraries.

Having juxtaposed the diagnostic PSVs into a condensed haplotype, we adapted bead-based single-molecule genotyping to haplotype the fused PCR product: (**Fig. 3** and **Supplementary Methods**) paramagnetic beads are coated with the amplified product of a single template molecule by performing PCR within an oil:water emulsion. These amplicon-coated beads are then genotyped by the addition of a single dye-labeled nucleotide (single-base extension) to a primer that is annealed directly flanking the variant site (**Fig. 3**). In previous work by other researchers, these beads were subsequently analyzed by flow cytometry[18]. We adapted this protocol to haplotyping by performing two successive rounds of PSV genotyping on beads immobilized within a polyacrylamide gel on a microscope slide and aligning the two images to assign haplotypes[26,27]. A universal primer tag on one of the outside primers of the haplotype-fusion PCR allows the amplified fusion product to be subsequently fused to beads coated with a matching double-biotinylated oligonucleotide. We demonstrated the successful application of this assay using single YACs and mixtures of YACs as templates for haplotype-fusion PCR (**Fig. 4**) before validating it on the two genomic DNAs from individuals with known inversion orientation (**Fig. 5a**). Finally, to demonstrate the utility of this assay, we genotyped thirteen previously uncharacterized individuals of diverse geographic origin. In twelve cases (92%) it was possible to unambiguously genotype this inversion event as both breakpoint haplotypes were diagnostic of either the inverted or

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

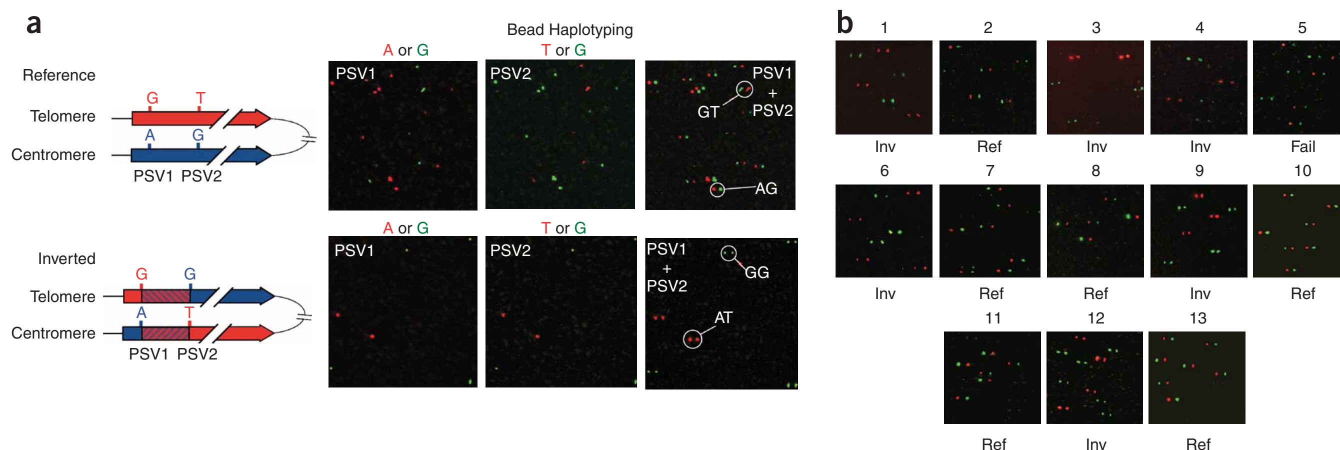

Figure 5 | Single-molecule haplotyping on genomic DNA. (**a**) The results of bead-based haplotyping using genomic DNA as the template for the initial haplotype-fusion PCR are shown for individuals known to have reference and inverted orientations of the Yp inversion. See **Figure 4** for details. (**b**) The results of bead haplotyping using genomic DNA as the template for the initial haplotype-fusion PCR are shown for 13 individuals of unknown orientations of the Yp inversion. Each image shows the results of overlaying the two genotyping images with the image of the second PSV shifted 10 pixels to the right. Inv, inversion; Ref, reference. Analysis for individual 5 failed inversion genotyping owing to an unexpected combination of breakpoint haplotypes (see text).

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

reference orientations (**Fig. 5b**). The single genotyping failure contained a mixture of two breakpoint haplotypes: one typical of the reference orientation and the other typical of the inverted orientation. This mixture of the two sets of expected breakpoint haplotypes most likely results from a gene conversion event that homogenizes PSV2.

DISCUSSION

In addition to genotyping known inversions and validating putative inversions, USF inversion genotyping can be used to screen panels of individuals for new inversion variants by targeting highly similar inverted repeats. Using USF, we have initiated a survey for inversion polymorphisms sponsored by known inverted repeats identified within databases of segmental duplications. In principle, the length of a haplotype that can be condensed by haplotype-fusion PCR is limited only by the size of the template DNA. The shearing effect of emulsion preparation is negligible on genomic DNA fragments 100–150 kb in length as judged by pulsed-field gel electrophoresis (**Supplementary Methods** and **Supplementary Fig. 2**), so it is reasonable to expect that template DNA molecules in emulsions can be at least 100 kb in length. The vast majority (97%) of inverted repeats in the segmental duplication databases[28] are less than 100 kb in length and so are amenable to USF genotyping. By contrast, long PCR methods targeted to unique sequences are typically restricted to inverted repeats 10 kb in length. Most duplicated sequences known to sponsor recurrent chromosomal rearrangements are greater than 10 kb in length[2].

In contrast to USF, PSVF inversion genotyping allows inversions with breakpoints within the longest inverted repeats to be genotyped, albeit with only a proportion of the inverted repeat assayed for inversion breakpoints. PSVF is likely to be most useful for genotyping known inversion events with well-characterized breakpoints as it requires that an inversion breakpoint fall within an interval defined by two known PSVs. To our knowledge, PSVs are not cataloged in the standard genome databases, although candidate PSVs can be identified from the human reference sequence by aligning duplicated sequences. Such PSVs are not necessarily fixed in the genome, so assays targeted to these variants may not work on all individuals. Gene conversion, the nonreciprocal transfer of sequence information from one homologous sequence to another, is one cause of PSVs not being fixed. Therefore it is important that any assay targeted towards PSVs is robust enough to distinguish between gene conversion events and inversion events. A single gene conversion event will generate an apparent recombinant haplotype at only one of the two inversion breakpoints. The PSVF inversion genotyping method described here haplotypes both inversion breakpoints: the combination of these two haplotypes allows the inversion event to be genotyped unambiguously. PSVF identifies discrepant pairs of breakpoint haplotypes that result from gene conversion events, without confusing them with either state of the inversion. We identified a single gene conversion event in this manner within the cohort subjected to blind PSVF inversion genotyping, demonstrating the importance of assaying both inversion breakpoints.

PSVF inversion genotyping could be used to survey for new inversions by tiling along a long (>150 kb) inverted repeat with overlapping assays, but using the present implementation this is likely to be overly laborious. The existence of hotspots of nonallelic homologous recombination within which most breakpoints fall is a common feature of rearrangements sponsored by duplicated sequences[29]. These hotspots are typically restricted to a small proportion of a much larger segmental duplication. PSVF inversion genotyping will be particularly useful for long inverted repeats that harbor hotspots for rearrangements. Moreover, the high-throughput identification and sequencing of fosmid clones harboring inversion breakpoints[8] should greatly increase the rate at which these hotspots are characterized.

How do our methods for inversion genotyping compare to alternative methods? First, in using a standard genomic DNA preparation for these assays, our methods are immediately applicable to a far greater range of samples than either cytogenetic or pulsed-field gel electrophoresis methods. Second, long-range PCR is limited by the processivity of DNA polymerases and is often not robust over distances of 10 kb or more, whereas haplotype-fusion

PCR is only limited by the size of the DNA fragments in the preparation. Our methods are thus applicable to all sizes of inverted repeats and not just the subset that can be PCR-amplified in their entirety. Finally, our assay methods are quick, robust and scalable to high-throughput (emulsions can be prepared in 96-well microtiter plates; data not shown). These features are all prerequisites for the population-scale genotyping required to elucidate the functional impact of any given inversion. In addition, as amplification only occurs in a small proportion (0.1% to 1%) of aqueous compartments within an emulsion, haplotype-fusion PCR should be readily amenable to multiplexing, using different sets of primers that do not compete for resources because amplification takes place in separate compartments.

It may be possible to genotype some inversions by typing nearby markers (for example, SNPs) in high linkage disequilibrium with the inversion[3,30], but this approach may be confounded if an inversion arises independently on different haplotype backgrounds. In the absence of population-scale genotyping studies little evidence has accrued for the recurrence or otherwise of neutral inversions[31]. It is clear, however, that pathogenic inversions can arise recurrently[9,10]. Furthermore, the dominant mutation process generating inversions (nonallelic homologous recombination) typically operates at rates of 10^{-3} to 10^{-5} per generation[32]. Thus many inversions are likely to arise recurrently, so the association between an inversion and flanking markers may be lower than for other variants.

The methods developed in this study have application beyond genotyping inversions. A logical extension of the methods is to the detection of recombinant haplotypes that are diagnostic of other types of chromosomal rearrangement (deletions, duplications and translocations) that arise either in the germline or somatic tissues. Also, these methods are applicable to any genome sequence in which duplicated sequences have been mapped. Assays that allow population-scale genotyping of all types of structural variation will greatly increase our understanding of the evolutionary history and medical consequences of this important class of genomic variation.

METHODS

DNA samples. Genomic DNA from males with known Y-chromosomal haplotypes were a kind gift of M. Jobling. YAC clones *692B1* and *821G7* come from the CEPH YAC library[33] ('reference'). YAC clones *AA4A7/11* and *AMY8* come from libraries constructed from CGM1 (ref. 34) and OXEN DNA[35], respectively ('inverted'). We obtained genomic DNA of individuals of diverse geographic ancestry from the European Collection of Cell Cultures (ECACC). We obtained genomic DNA from hemophilia A patients as part of the UK National Haemophilia A Mutation Database project and the intron 22 inversion previously typed using the long PCR method[12,13]. All samples had been collected with patients' informed written consent.

Quantifying the template:compartment ratio. Water-in-oil emulsions prepared as described below contain spherical aqueous compartments with an average diameter of 15 μm[19]. An emulsion made from 200 μl oil and 100 μl aqueous phase is therefore expected to contain 5.7×10^7 aqueous compartments. The target region of the fusion PCRs lie on the X or Y chromosomes of male individuals, so emulsions made using 200 ng genomic DNA contain approximately 30,000 amplifiable template molecules. Consequently only 1 in around 1,900 aqueous compartments contains a suitable template molecule for fusion PCR, and so amplification derives from a single molecule in the overwhelming majority of cases.

Haplotype-fusion PCR: Yp inversion. We prepared PCRs in a total volume of 100 μl, containing 1× Phusion HF buffer (Finnzymes), an additional 1 mM $MgCl_2$, 250 μM dNTPs, 1 μM primers InvF1 (Sigma; 5′-TGCCTGAGGAAGCCCTGTAGTTAG-3′) and InvR2 (Sigma; 5′-GGAGCACGCTATCCCGTTAGACCCC TCCATTTTAGAGACAGGTGC-3′), 10 nM primer InvFusionR1 (Sigma; 5′-GATATATAAATGTTGCAATGGGGGAGCATGAGCTG CAGGAGGCCTTTGC-3′), 200 ng (6×10^4 copies) genomic DNA, 16 units Phusion DNA polymerase (Finnzymes) and added this aqueous phase dropwise over ~50 s to 200 μl oil phase (4.5% vol/vol Span 80, 0.4% vol/vol Tween 80 and 0.05% Triton X-100 dissolved in light mineral oil (all from Sigma)), in a 2 ml Cryo-Vial (Corning) while stirring with a magnetic bar (8 × 3 mm with a pivot ring; VWR) at 1,000 r.p.m.[19]. We continued stirring for a total of 5 min after which we overlaid the emulsions with 30 μl of mineral oil (Sigma) before thermocycling (98 °C for 30 s; 33 cycles at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 15 s; 72 °C for 5 min).

We then disrupted emulsions by the addition of hexane (Aldrich) and reamplified 2 μl of each clean PCR product in a total volume of 50 μl using primers InvF1 an InvR2 (1× Phusion HF buffer, 200 μM dNTPs, 300 nM primers and 1 unit Phusion DNA polymerase) under the same cycling conditions. We used ethanol precipitation to concentrate PCR products, and resuspended them in 18 μl 1× Phusion HF buffer, to which we added 1.6 units Proteinase K (Sigma). We incubated reactions at 56 °C for 1 h to digest the polymerase and 95 °C for 10 min to denature the Proteinase K, after which we used gel electrophoresis to quantify the DNA.

Haplotype-fusion PCR: hemophilia inversion. We used a similar protocol to that for the Y inversion, but with 1 μM primers F1 (Sigma; 5′-ATTTACCTCCCCTGCCATGGAAATC-3′), R2 (Sigma; 5′-ACAATCTGGAGGTGAATGCTTAGTTC-3′) and R3 (Sigma; 5′-CATTTAACCAAAAAGCACTTGTAAGCCA-3′), and 10 nM primer R1 (Sigma; 5′-CCAGAATTTGGCCCATAGCCTGCTGTGT TAAGCATTGAGACAACACCAAC-3′). Thermocycling conditions were as with the Y inversion, but using 40 cycles and a 66 °C annealing step. After disruption of the emulsion by hexane, we reamplified fused PCR products in 50 μl of aqueous solution containing 1× JumpStart polymerase PCR buffer (Sigma; 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (wt/vol) gelatin), 200 μM dNTPs, 600 nM primer F1, 300 nM primers R2 and R3, 1 μl 100-fold–diluted PCR product and 1.25 units JumpStart polymerase (Sigma). Cycling conditions were: 94 °C for 2 min, 35 cycles at 94 °C for 30 s, 62 °C for 30 s and 72 °C for 15 s; 72 °C for 10 min). We electrophoresed 15 μl of each product of this second amplification in a 2.5% agarose gel alongside 5 μl HyperLadder IV Quantitative DNA marker (Bioline) and visualized gels by ethidium bromide staining under ultraviolet light.

Additional Methods. Details of bead-based PCR in emulsion and pulsed-field gel electrophoresis are available in **Supplementary Methods**.

© 2006 Nature Publishing Group http://www.nature.com/naturemethods

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS
This work was funded by the Wellcome Trust. We thank M. Jobling for the kind gift of DNA, M. Ross for insights into Xq28, J. McCafferty and W. Howat for help with imaging, J. Collins and A. Coffey for guidance on pulsed-field gel elecrophoresis, and P. Holliger and Z. Oliynyk for advice on preparation of emulsions.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hogervorst, F.B. *et al.* Large genomic deletions and duplications in the *BRCA1* gene identified by a novel quantitative method. *Cancer Res.* **63**, 1449–1453 (2003).
2. Stankiewicz, P. & Lupski, J.R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
3. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
4. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
5. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
6. Sharp, A.J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
7. Armour, J.A., Barton, D.E., Cockburn, D.J. & Taylor, G.R. The detection of large deletions or duplications in genomic DNA. *Hum. Mutat.* **20**, 325–337 (2002).
8. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
9. Bunge, S. *et al.* Homologous nonallelic recombinations between the iduronate-sulfatase gene and pseudogene cause various intragenic deletions and inversions in patients with mucopolysaccharidosis type II. *Eur. J. Hum. Genet.* **6**, 492–500 (1998).
10. Rossiter, J.P. *et al.* Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. *Hum. Mol. Genet.* **3**, 1035–1039 (1994).
11. Osborne, L.R. *et al.* A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat. Genet.* **29**, 321–325 (2001).
12. Liu, Q., Nozari, G. & Sommer, S.S. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. *Blood* **92**, 1458–1459 (1998).
13. Small, K., Iber, J. & Warren, S.T. Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat. Genet.* **16**, 96–99 (1997).
14. Antonarakis, S.E. *et al.* Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* **86**, 2206–2212 (1995).
15. Jobling, M.A. *et al.* A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.* **8**, 1391–1394 (1998).
16. Wetmur, J.G. *et al.* Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes. *Nucleic Acids Res.* **33**, 2615–2619 (2005).
17. Yon, J. & Fried, M. Precise gene fusion by PCR. *Nucleic Acids Res.* **17**, 4895 (1989).
18. Dressman, D. *et al.* Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **100**, 8817–8822 (2003).
19. Ghadessy, F.J., Ong, J.L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA* **98**, 4552–4557 (2001).
20. Bagnall, R.D., Giannelli, F. & Green, P.M. Polymorphism and hemophilia A causing inversions in distal Xq28: a complex picture. *J. Thromb. Haemost.* **3**, 2598–2599 (2005).
21. Ross, M.T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
22. Vogt, P.H. *et al.* Report of the third international workshop on Y chromosome mapping 1997. *Cytogenet. Cell Genet.* **79**, 2–16 (1997).
23. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
24. Williams, G. *Mapping Studies of the Centromeric Region of the Human Y Chromosome.* D. Phil. Thesis. University of Oxford, UK (1998).
25. Cooper, K.F., Fisher, R.B. & Tyler-Smith, C. Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y-chromosome. *J. Mol. Biol.* **230**, 787–799 (1993).
26. Mitra, R.D. *et al.* Digital genotyping and haplotyping with polymerase colonies. *Proc. Natl. Acad. Sci. USA* **100**, 5926–5931 (2003).
27. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
28. She, X. *et al.* Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**, 927–930 (2004).
29. Lupski, J.R. Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol.* **5**, 242 (2004).
30. Feuk, L. *et al.* Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet.* **1**, e56 (2005).
31. Broman, K.W. *et al.* Common long human inversion polymorphism on chromosome 8p'. In *Science and Statistics: A Festschrift for Terry Speed* (ed, Goldstein, D.R.) 237–245 (IMS Lecture Notes - Monograph Series, Beechwood, Ohio, 2003).
32. Shaffer, L.G. & Lupski, J.R. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu. Rev. Genet.* **34**, 297–329 (2000).
33. Cohen, D., Chumakov, I. & Weissenbach, J. A first-generation physical map of the human genome. *Nature* **366**, 698–701 (1993).
34. Brownstein, B.H. *et al.* Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348–1351 (1989).
35. Monaco, A.P. *et al.* Isolation of the human sex determining region from a Y-enriched yeast artificial chromosome library. *Genomics* **11**, 1049–1053 (1991).

© 2006 Nature Publishing Group http://www.nature.com/naturemethods